

PATENT ABSTRACTS OF JAPAN

(11)Publication number : **07-239843**

(43)Date of publication of application : **12.09.1995**

(51)Int.Cl.

G06F 17/16

G06F 15/16

(21)Application number : **06-028110**

(71)Applicant : **SONY CORP**

(22)Date of filing : **25.02.1994**

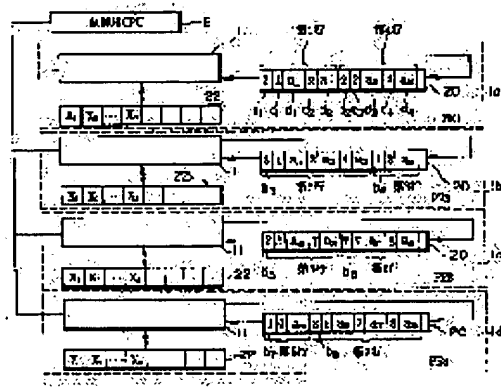
(72)Inventor : **KATO YASUNOBU**

(54) PARALLEL ARITHMETIC PROCESSORS

(57)Abstract:

PURPOSE: To reduce the storage capacity required for arithmetic and to carry out more efficient and higher-speed arithmetic.

CONSTITUTION: A control part 2 divides a matrix A for each row, forms arithmetic data for each row from the combination of element number data b_i expressing the number of non-zero elements in one row, column numbers c_i concerning all the non-zero elements and values d_i of elements and supplies the arithmetic data and data x_i of a vector X to respective arithmetic processing parts 1a-1d, and the respective arithmetic processing parts 1a-1d store the arithmetic data in a SAM 20 and store the data x_i of the vector X in a RAM 22. The respective arithmetic processing parts 1a-1d read the element number data b_i stored in the SAM 20, read the column numbers c_i and the values d_i of elements based on the element number data b_i and read the data x_i of the vector X corresponding to the column numbers c_i from the RAM 22, arithmetic is performed by an arithmetic



part 11, and the arithmetic to calculate the inner product of the matrix A and the vector X is parallelly executed at the arithmetic parts 1a-1d.

LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

*** NOTICES ***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The storage section which memorizes operation data, and two or more data-processing sections respectively equipped with the operation part which calculates to the operation data memorized by this storage section, The number data of elements which divide data into a field and express the number of the non-zero elements which are not 0 in this field, Operation data are formed from the group of the location data showing the location in the above-mentioned field of the above-mentioned non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data. Supply these operation data to each data-processing section, and it has the control section which controls the operation of each data-processing section. Each above-mentioned data-processing section the above-mentioned number data of elements memorized in each above-mentioned storage section Read-out, The parallel operation processor characterized by reading the above-mentioned location data and value data based on these number data of elements, calculating by operation part, and performing the operation to data to juxtaposition in two or more above-mentioned data-processing sections.

[Claim 2] The above-mentioned control section is a parallel operation processor according to claim 1 which divides the data of the letter of a matrix for every line, uses the number of non-zero elements in this one line as the above-mentioned number data of elements, and is characterized by to form the above-mentioned operation data per line of the data of the letter of a matrix, and to supply each data-processing section by using the location within the one above-mentioned line of a non-zero element as the above-mentioned location data.

[Claim 3] The above-mentioned storage section is a parallel operation processor according to claim 1 or 2 characterized by having the serial access memory which performs the continuous writing and continuous read-out of data.

[Claim 4] The above-mentioned control section is a parallel operation processor given in any 1 term of claim 1 characterized by supplying operation data to each data-processing section so that the load of each data-processing section may become equal thru/or claim 3.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to the parallel operation processor which calculates the data of the letter of a matrix especially about the parallel operation processor which calculates data to juxtaposition by two or more operation part.

[0002]

[Description of the Prior Art] By the development and spread of computers in recent years, the large-scale scientific calculation currently considered for implementation to be before impossible is also realistic. For example, with the finite difference method and the finite element method which are used for electromagnetic-field analysis or the flow analysis of a fluid, and a boundary element method, the so-called matrix operation, such as very big characteristic value count of the solution of the simultaneous equation of size or a matrix, appears frequently in the course of those analyses.

[0003] Moreover, in order to raise analysis precision, it is possible to enlarge a matrix, but if a matrix becomes large, even if it carries out matrix operation with the supercomputer in which a high-speed operation is possible, time amount borrows it very much, and high-speed processing is desired.

[0004] Although storage with large storage capacity was called for with improvement in the speed of the processing speed of the data-processing section in the matrix operation of the above big sizes at high speed, since the storage capacity and the working speed of storage had an opposite relation, generally the storage with large storage capacity had the problem from which equipment will become big very at an expensive price at high speed.

[0005] for this reason -- for example, although it is high-speed between the data-processing section (CPU) 61 and the mass main storage 63 as shown in drawing 7, it has the cache memory 62 with a comparatively small capacity, data are beforehand read from main storage 63 to cache memory 62 through the bus line 65, when CPU61 reads data from cache memory 62 through a bus line 64, high-speed read-out of data is realized and high speed ***** is used in the operation.

[0006] Moreover, as shown in drawing 8, in order to perform the same operation to a series of data, such as matrix operation, at a high speed, [for example,] A vector 77, i.e., the vector register which memorizes a series of data, two or more preparations, It calculates to a series of data memorized by this vector register 77 with the computing element of the adder 70 of the floating point (FP:FloatingPoint), a multiplier 71, and divider 72 grade. The processing unit (the so-called supercomputer) which accelerated the operation is known by pipelining these processings.

[0007] In this processing unit, data are transmitted to the above-mentioned vector register 77 through data lines 78 and 80 with vector I/O device (vector I/O) 79 in advance of an operation from main storage 81. And the data transmitted to the vector register 77 are again written in the vector register 77 through the output bus line 74, after each above-mentioned computing elements 70-72 are supplied through the input bus line 73 and an operation is completed. Furthermore, the data written in the vector register 77 which is the result of an operation are again memorized by main storage 81 through vector I/O79.

[0008]

[Problem(s) to be Solved by the Invention] However, if data other than the data which read data from main storage 63 to cache memory 62 beforehand are needed, it is necessary to read the processing unit which used the above-mentioned cache memory 62 from main storage 63 (if a mistake hit is carried out). Generally, by the operation of the big matrix of size, the capacity of cache memory 62 was small, the data which have not gone into cache memory 62 needed to be read from main storage 63 each time, as for the inside of the time amount which is this read-out, CPU61 stopped and there was a problem to which the operation speed as the whole falls.

[0009] Moreover, also in the processing unit equipped with the above-mentioned vector register

77, since it was necessary to read from main storage 81 by the operation of the big matrix of size about the element (data) of the matrix which is not memorizable to the vector register 77 since the capacity of the vector register 77 is small, and it was generally necessary to calculate, operation speed was reduced.

[0010] Then, paying attention to the descriptions, such as read-out sequence of the data in the case of matrix operation, as shown in drawing 9 and drawing 10, this applicant had the serial access memory which performs the continuous writing and continuous read-out of data in each data-processing section of a parallel operation processor which calculates to juxtaposition in two or more data-processing sections, accelerated operation speed, and has proposed previously the parallel operation processor which realized low-cost-izing and a miniaturization as Japanese Patent Application No. No. 149597 [five to].

[0011] The m data-processing sections 101a and 101b which calculate as this parallel operation processor is shown in drawing 9, and 101 c*****101m, The control section 102 which outputs a control signal, operation data, etc. which control a these data-processing sections [101a-101m] operation, It has the data bus 103 which supplies the operation data from this control section 102 etc. to the above-mentioned data-processing sections 101a-101m, and the control bus 104 which supplies the control signal from the above-mentioned control section 102 to the above-mentioned data-processing sections 101a-101m.

[0012] Moreover, each above-mentioned data-processing sections 101a-101m The storage section 110 which memorizes the operation data supplied from the above-mentioned control section 102 as shown in drawing 10, The operation part 111 which calculates to the data memorized by this storage section 110, The buffer 113 which performs the communication link with the above-mentioned data bus 103 through an internal data bus 112, The buffer 115 which performs the communication link with the above-mentioned control bus 104 through the internal control bus 114, and a control signal required for a communication link are generated, and it has the control signal generating section 116 supplied to a buffer 115 and operation part 111.

[0013] Furthermore, the above-mentioned storage section 110 consists of mass continuation I/O memory (henceforth mass serial access memory (SAM)) 120 which performs the continuous writing and continuous read-out of data, and high-speed general-purpose RAM122 of capacity which performs the random writing and random read-out of the small capacity SAM 121 with a small capacity, and data and which is not so large, as shown in above-mentioned drawing 10.

[0014] In this parallel operation processor, when asking for the product of the matrix A shown, for example in the following formula 1, and Vector X, a control section 102 divides Matrix A, and assigns and supplies the data for every line to each data-processing section. For example, in the case of a formula 1, the data for two lines of 8x8 matrices are assigned to the four data-processing sections 101a-101d, and each data-processing sections 101a-101d are supplied. For example, data of the 1st line and the 5th line are supplied to data-processing section 101a, data of the 2nd line and the 6th line are supplied to data-processing section 101b, data of the 3rd line and the 7th line are supplied to data-processing section 101c, and data of the 4th line and the 8th line are supplied to 101d of data-processing sections.

[0015]

[Equation 1]

[0016] And a control section 102 is the element x_i of Vector X. By supplying each data-processing sections 101a-101d, each data-processing sections 101a-101d store each line of Matrix A in large capacity SAM 120, as shown in drawing 11, and it is the element x_i of Vector X. It stores in RAM122. And it is the element x_i of the vector X continuously corresponding to [to the time of an operation / in the each data-processing sections / 101a-101d / operation part 111] read-out and Element a_{ij} for the element a_{ij} of each line. Element y_i of the vector Y as a result of performing the operation shown in read-out and the following formula 2 from RAM122 and being based on this operation It memorizes to RAM122. And a control section 102 is the element y_i of Vector Y called for in each data-processing sections 101a-101d. Vector Y was searched for unitedly. Consequently, this parallel operation processor can perform matrix operation now at a high speed.

[0017]

[Equation 2]

[0018] On the other hand, the large-scale matrix which appears in scientific calculation was a sparse matrix most whose elements are 0 in many cases, and if all the elements of this sparse matrix were memorized at the time of an operation, there was a problem which needs huge storage capacity, so that it might be represented by a finite difference method, the finite element method, linear programming, electric circuit analysis, etc.

[0019] Then, the processing unit which shortens the operation time was considered by not performing the operation about zero element whose value of the element in this sparse matrix is 0, or not memorizing zero element.

[0020] however, in the processing unit which does not perform the operation about zero element of the above-mentioned sparse matrix the need of the positional information of the data which are not 0 in the sparse matrix used as the candidate for an operation being needed on the occasion of an operation, memorizing all elements, and judging at the time of an operation -- or It is difficult to memorize the positional information of the data which are not 0 in a sparse matrix beforehand, and to apply to the parallel operation processor which the whole operation time was increased as a result, and used the above-mentioned serial access memory by the read time of such information.

[0021] Moreover, in the processing unit which does not memorize zero element of the above-mentioned sparse matrix, there was a problem on which it is necessary to memorize the positional information of operation data in addition to operation data, the storage capacity which an operation takes increases, and the operation time as the whole increases by the read time of positional information.

[0022] This invention can be made in view of the above troubles, can calculate the matrix of big size at a high speed, can reduce the storage capacity which an operation takes, can raise cost

performance, and aims at offer of the parallel operation processor which made the miniaturization possible.

[0023]

[Means for Solving the Problem] In order to solve an above-mentioned technical problem, the parallel operation processor concerning this invention The storage section which memorizes operation data, and two or more data-processing sections respectively equipped with the operation part which calculates to the operation data memorized by the storage section, The number data of elements which divide data into a field and express the number of the non-zero elements which are not 0 in a field, Operation data are formed from the group of the location data showing the location in the field of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data. Supply operation data to each data-processing section, and it has the control section which controls the operation of each data-processing section. It is characterized by for each data-processing section reading location data and value data based on read-out and the number data of elements, calculating the number data of elements memorized in each storage section by operation part, and performing the operation to data to juxtaposition in two or more data-processing sections.

[0024] Moreover, it is characterized by for a control section dividing the data of the letter of a matrix for every line, and for the parallel operation processor concerning this invention using the number of non-zero elements in one line as the number data of elements, forming operation data per line of the data of the letter of a matrix by using the location within one line of a non-zero element as location data, and supplying them to each data-processing section.

[0025] Moreover, the parallel operation processor concerning this invention is characterized by equipping the storage section with the serial access memory which performs the continuous writing and continuous read-out of data.

[0026] Moreover, the parallel operation processor concerning this invention is characterized by a control section supplying operation data to each data-processing section so that the load of each data-processing section may become equal.

[0027]

[Function] A control section divides the data of the letter of a matrix into a field, forms operation data from the group of the number data of elements showing the number of the non-zero elements which are not 0 in a field, the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data, supplies operation data to each data-processing section, and controls the operation of each data-processing section by the parallel-operation processor concerning this invention. Each data-processing section supplies the operation data supplied from the control section to the storage section, and operation data are memorized by the storage section. And each data-processing section reads location data and value data based on read-out and the number data of elements, and calculates to juxtaposition the number data of elements memorized by the storage section.

[0028] In the parallel operation processor concerning this invention, moreover, a control section Divide the data of the letter of a matrix for every line, and the number of non-zero elements in one line is used as the above-mentioned number data of elements. The location within the one above-mentioned line of a non-zero element is used as the above-mentioned location data, the above-mentioned operation data are formed per line of the data of the letter of a matrix from the group of the value data showing the value of the non-zero element of the location corresponding to a series of location data, operation data are supplied to the data-processing section, and the

operation of each data-processing section is controlled. Each data-processing section supplies the operation data supplied from the control section to the storage section, and operation data are memorized by the storage section. And each data-processing section reads location data and value data based on read-out and the number data of elements, and calculates to juxtaposition the number data of elements memorized in the storage section.

[0029] Moreover, a control section divides data into a field, forms operation data from the group of the number data of elements showing the number of the non-zero elements which are not 0 in a field, the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data, supplies operation data to each data-processing section, and controls the operation of each data-processing section by the parallel-operation processor concerning this invention. Each data-processing section supplies continuously the operation data supplied from the control section to serial access memory, and operation data are memorized by serial access memory. And each data-processing section reads the number data of elements memorized to serial access memory, reads location data and value data continuously based on the number data of elements, and calculates to juxtaposition.

[0030] In the parallel operation processor concerning this invention, moreover, a control section The number data of elements which divide data into a field and express the number of the non-zero elements which are not 0 in a field, Operation data are formed from the group of the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data, operation data are supplied to each data-processing section so that the load of each data-processing section may become equal, and the operation of each data-processing section is controlled. Each data-processing section supplies the operation data supplied from the control section to the storage section, and operation data are memorized by the storage section. And from the storage section, each data-processing section reads location data and value data based on read-out and the number data of elements, calculates by operation part, and calculates the number data of elements to juxtaposition.

[0031]

[Example] Hereafter, the suitable example of the parallel operation processor concerning this invention is explained to a detail, referring to a drawing.

[0032] This parallel operation processor For example, the m data-processing sections 1a and 1b which calculate as shown in drawing 1 and 1 c*****1m, The control section 2 which outputs a control signal, operation data, etc. which control these operation control sections 1a-1m, It has the data bus 3 which supplies the operation data from this control section 2 to the above-mentioned operation control sections 1a-1m, and the control bus 4 which supplies the control signal from the above-mentioned control section 2 to the above-mentioned operation control sections 1a-1m.

[0033] Each above-mentioned operation control sections 1a-1m For example, the storage section 10 which memorizes the operation data supplied from the above-mentioned control section 2 as shown in drawing 2 , The operation part 11 which calculates to the data memorized by this storage section 10, The buffer 13 which performs the communication link with the above-mentioned data bus 3 through an internal data bus 12, the buffer 15 which performs the communication link with the above-mentioned control bus 4 through the internal control bus 14, and a control signal required for a communication link are generated, and it has the control signal generating section 16 supplied to a buffer 15 and operation part 11. Each data-processing

sections 1a-1m calculate according to the control signal from a control section 2.

[0034] And the above-mentioned storage section consists of mass continuation I/O memory (henceforth mass serial access memory (SAM)) 20 which performs the continuous writing and continuous read-out of data, and high-speed general-purpose RAM22 of capacity which performs the random writing and random read-out of the small capacity SAM 21 with a small capacity, and data and which is not so large, as shown in above-mentioned drawing 2.

[0035] Moreover, above-mentioned size ** SAM 20 and the small capacity SAM 21 For example, the memory cell array 40 to which the writing and read-out of data are performed as shown in drawing 3, The read-out line address counter 41 which generates the read-out address of the data to the line of this memory cell array 40, The write-in line address counter 42 which generates the write-in address of the data to the line of the above-mentioned memory cell array 40, The read-out train address counter 43 which generates the read-out address of the data to the train of the above-mentioned memory cell array 40, It has the write-in train address counter 44 which generates the write-in address of the data to the train of the above-mentioned memory cell array 40, data input Rhine 45 which inputs input data Din, and data output Rhine 46 which outputs output-data Dout.

[0036] And at the time of the writing of data, the reset light signal RSTW is supplied to the above-mentioned write-in line address counter 42 and the above-mentioned write-in address train counter 44 with the write enable signal WE, a counter value is reset, the write-in clock WCK supplied henceforth is counted, and the address which writes in is generated. Furthermore, with the above-mentioned write-in clock WCK, the write-in data Din are continuously supplied from data input Rhine 45, and are memorized by the memory cell 40.

[0037] Moreover, at the time of read-out of data, the reset lead signal RSTR is supplied to the above-mentioned read-out line address counter 41 and the above-mentioned read-out address train counter 43 with the lead enable signal RE, a counter value is reset, the read-out clock RCK supplied henceforth is counted, and the address which performs read-out is generated. Furthermore, with the above-mentioned read-out clock RCK, the read-out data Din are continuously read from a memory cell 40, and it is outputted to data output Rhine 46.

[0038] Next the operation which asks this parallel operation processor actuation for the product of a matrix and a vector is made into an example, and it explains. Generally, the product Y of the matrix A as shown in the following formulas 3, and Vector X is the element yi of the product vector Y by the operation as shown in the following formula 4. It asks by asking.

[0039]

[Equation 3]

[0040]

[Equation 4]

[0041] Here, as Matrix A shows the operation of the above-mentioned formula 3 in the following formulas 5, when the element of the great portion of matrix is the so-called sparse matrix which is 0, it can ask for the product Y of Matrix A and Vector X by the operation as shown in the following formula 6.1 - a formula 6.8.

[0042]

[Equation 5]

[0043]

[Equation 6]

[0044] So, with this parallel operation equipment, a control section 2 divides Matrix A into each line, forms operation data from the group of the number data of elements showing the number of the non-zero elements in one line, the location data (row number) about all non-zero elements, and value data (value of an element), and supplies this operation data to each data-processing sections 1a-1m. In the case of the above-mentioned formula 5, since the non-example elements of the 1st line are a11 and a12, the number data of elements are set to "2", and, specifically, location data are set to "1" and "2", respectively, for example. Moreover, since the non-example elements of the 2nd line are a21, a22, and a24, the number data of elements are set to "3", location data are set to "1", "2", and "4", respectively, and value data are set to "a21", "a22", and "a24", respectively, for example.

[0045] Moreover, a control section 2 compares the number of the non-zero elements of each line, and it assigns each data-processing sections [1a-1m] operation data so that an each data-processing sections [1a-1m] load may become equal. Allocation of concrete operation data memorizes the number of the non-zero elements of each line of for example, the matrix A, and is performed by choosing the line assigned to each data-processing sections 1a-1m so that the number of these non-zero elements may become equal. For example, in the case of the above-mentioned formula 5, since the number of the non-zero elements from the 1st line of Matrix A to the 8th line is 2, 3, 1, 2, 2, 1, 3, and 2, respectively, as shown, for example in drawing 4 In the case where it has the four data-processing sections 1a, 1b, 1c, and 1d Data of the 1st line and the 4th line are supplied to data-processing section 1a, data of the 2nd line and the 3rd line are supplied to data-processing section 1b, data of the 5th line and the 8th line are supplied to data-processing section 1c, and data of the 6th line and the 7th line are supplied to 1d of data-processing sections.

[0046] Consequently, the number of the non-zero elements which the operation data for two lines are respectively supplied to each data-processing sections 1a-1d, and are supplied to each data-processing sections 1a-1d is four pieces respectively, and the each data-processing sections [1a-1d] load (operation) is equal.

[0047] And the each processing units [1a-1d] operation part 11 calculates by reading the above-mentioned operation data (the number data of elements, location data, value data) memorized to large capacity SAM 20 according to the control signal from a control section 2.

[0048] the next -- each data-processing section 1 -- the case where the product of the matrix A which shows actuation of the actual operation data in a-1d in the above-mentioned formula 3, and Vector X is calculated is explained to an example using the flow chart shown in 5 Fig. First, as operation data are supplied to each data-processing sections 1a-1d in advance of the operation, for example, it is shown in drawing 4 , the operation data for every line of Matrix A are memorized by large capacity SAM 20, and the data of Vector X are memorized by high-speed general-purpose RAM22. Specifically to the large capacity SAM 20 of data-processing section 1a The number data b1 of elements showing the number of the non-zero elements of the 1st line of Matrix A (2), The number data b2 of elements which 2 sets of location data cp (1 p=2) and the value data dp (1 11 a 2a 12) are memorized continuously, then express the number of the

non-zero elements of the 4th line of Matrix A (2), 2 sets of location data cp (2 p= 4) and the value data dp (2 42 a 4a 44) are memorized. Furthermore, the data xq (q= 1, 2...8) of Vector X are memorized by high-speed general-purpose RAM22. Moreover, although not illustrated, the data in which the line count (the line count in their duty) and line number (i) which were assigned to each data-processing sections 1a-1d with operation data are shown are supplied to each data-processing sections 1a-1d.

[0049] And the operation data memorized by large capacity SAM 20 as mentioned above are continuously read in the sequence memorized at the time of an operation. For example, at above-mentioned data-processing section 1a, it is the number data b1 of elements of the 1st line of Matrix A first. (2) is read. Then, 2 sets of location data cp The value data dp (1 11 a 2a 12) are read continuously, and it is the number data d2 of elements of the 4th line of Matrix A further. (2) and 2 sets of location data cp The value data dp (2 42 a 4a 44) are read.

[0050] And a control section 2 supplies the control signal which starts an operation to each data-processing sections 1a-1d through a control bus 4 and the internal control bus 14, and progresses to step S1. And in step S1, operation part 11 sets to 1 the count variable k which counts a processing line count, and progresses to step S2.

[0051] In step S2, operation part 11 compares the value and the line count in its duty of the count variable k, if the count variable k is below the line count in its duty, it will progress to step S3, and if the count variable k is size from the line count in its duty, since processing for the line count in its duty is already completed, it will be ended.

[0052] In step S3, operation part 11 reads from large capacity SAM 20, the several m b, i.e., above-mentioned number data of elements, of a non-zero element, and progresses to step S4.

[0053] In step S4, the value of the count variable L is set to 1, it sets the value of Variable yi to 0, and operation part 11 progresses to step S5.

[0054] In step S5, operation part 11 measures several m of the value of the count variable L, and a non-zero element, if the value of the count variable L is several m or less of a non-zero element, it will progress to step S6, and if the value of the count variable L is size from several m of a non-zero element, since processing for one line is already completed, it will progress to step 11.

[0055] It sets to step S6 and operation part 11 is the row number cp of a non-zero element (j), i.e., above-mentioned location data, from large capacity SAM 20. It reads and progresses to step S7.

[0056] Setting to step S7, operation part 11 is, the value (aij) dp, i.e., the above-mentioned value data, of an element of the matrix from large capacity SAM 20. It reads and progresses to step S8.

[0057] It is the value xj of the element of the vector X on step S8 and corresponding to the row number (j) of high-speed general-purpose RAM22 to the above-mentioned non-zero element in operation part 11. It reads and progresses to step S9.

[0058] Setting to step S9, operation part 11 is Variable yi. The value (aij) of the element of a matrix, and value xj of the element of Vector X A product is added and it progresses to step 10.

[0059] In step S10, operation part 11 adds 1 to the value of the count variable L, and returns to step S5. That is, processing from the above-mentioned step S5 to step S10 is repeated about the data for one line, and they are the value (aij) of the element for one line, and the value xj of the element of Vector X. It is Variable yi about a product. It will add.

[0060] After processing for one line is completed, the value of the count variable L serves as size from several m of a non-zero element, and it progresses to step S11 from step S5, and sets to step S11. And operation part 11 Variable yi called for as mentioned above Value yi, i.e., the value of

the element of the product vector Y, In step S12 which writes in and follows high-speed general-purpose RAM22, 1 is added to the value of the count variable k, and it returns to step S2. That is, processing from step S2 to step S12 is repeated until processing for the line count in its duty is completed.

[0061] As mentioned above, each data-processing sections 1a-1d are the values y_i of the assigned element of the product vector Y according to the above-mentioned formula 6.1 - a formula 6.8. As it calculates, for example, is shown in drawing 6, it memorizes to high-speed general-purpose RAM22. And value y_i of the element of this product vector Y It is read by the control signal from a control section 2, and a control section 2 is the value y_i of the element of the product vector Y from each data-processing sections 1a-1d. An operation is ended unitedly.

[0062] Since this invention was applied to the parallel operation processor which performs the operation which asks for the product of a matrix and a vector, while a value cannot perform the operation to the zero element which is 0 but can accelerate an operation, in this parallel operation processor, the storage capacity of the storage section which an operation takes can be reduced, so that clearly from the above explanation. Moreover, since operation data were assigned to each data-processing section so that the load of each data-processing section might become equal as mentioned above, operation effectiveness can be raised and the whole operation time can be shortened.

[0063] In addition, the technical thought of this invention is not limited to an above-mentioned example, and can also apply the operation to process also to the operation which asks for an above-mentioned matrix, not only the product of a vector but a matrix, and the product of matrices. Moreover, the field which divides data can also perform the operation which asks for a matrix and the product of matrices at a high speed by dividing one matrix for every line, dividing the matrix of another side for every train, and forming operation data by the operation which asks not only for the line unit of an above-mentioned matrix but for an above-mentioned matrix and the above-mentioned product of matrices. Moreover, it is clear that it is applicable also to the operation to the data with which the data which serve as a candidate for an operation, for example are not restricted to an above-mentioned matrix, either, and a single string continues.

[0064]

[Effect of the Invention] In the parallel operation processor applied to this invention by above-mentioned explanation so that clearly The number data of elements which a control section divides the data of the letter of a matrix into a field, and express the number of the non-zero elements which are not 0 in a field, Operation data are formed from the group of the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data. Operation data are supplied to each data-processing section, and the operation of each data-processing section is controlled. Each data-processing section The number data of elements with which each operation part was memorized by each storage section by reading location data and value data based on read-out and the number data of elements, and calculating to juxtaposition by supplying the operation data supplied from the control section to the storage section It can calculate at a high speed and the storage capacity which an operation takes can be reduced.

[0065] Moreover, in the parallel operation processor concerning this invention, by dividing for every line of the data of the letter of a matrix, forming operation data, and calculating to juxtaposition in two or more data-processing sections, a matrix can be calculated at a high speed and the storage capacity which an operation takes can be reduced.

[0066] Moreover, in the parallel operation processor concerning this invention, since the storage

section was equipped with the serial access memory which performs the continuous writing and continuous read-out of data, cost performance can be raised and equipment can be miniaturized. [0067] Moreover, in the parallel operation processor concerning this invention, a more efficient and high-speed operation can be performed by supplying operation data to each data-processing section so that the load of each data-processing section may become equal.

TECHNICAL FIELD

[Industrial Application] This invention relates to the parallel operation processor which calculates the data of the letter of a matrix especially about the parallel operation processor which calculates data to juxtaposition by two or more operation part.

PRIOR ART

[Description of the Prior Art] By the development and spread of computers in recent years, the large-scale scientific calculation currently considered for implementation to be before impossible is also realistic. For example, with the finite difference method and the finite element method which are used for electromagnetic-field analysis or the flow analysis of a fluid, and a boundary element method, the so-called matrix operation, such as very big characteristic value count of the solution of the simultaneous equation of size or a matrix, appears frequently in the course of those analyses.

[0003] Moreover, in order to raise analysis precision, it is possible to enlarge a matrix, but if a matrix becomes large, even if it carries out matrix operation with the supercomputer in which a high-speed operation is possible, time amount borrows it very much, and high-speed processing is desired.

[0004] Although storage with large storage capacity was called for with improvement in the speed of the processing speed of the data-processing section in the matrix operation of the above big sizes at high speed, since the storage capacity and the working speed of storage had an opposite relation, generally the storage with large storage capacity had the problem from which equipment will become big very at an expensive price at high speed.

[0005] for this reason -- for example, although it is high-speed between the data-processing section (CPU) 61 and the mass main storage 63 as shown in drawing 7 , it has the cache memory 62 with a comparatively small capacity, data are beforehand read from main storage 63 to cache memory 62 through the bus line 65, when CPU61 reads data from cache memory 62 through a bus line 64, high-speed read-out of data is realized and high speed ***** is used in the operation.

[0006] Moreover, as shown in drawing 8 , in order to perform the same operation to a series of data, such as matrix operation, at a high speed, [for example,] A vector 77, i.e., the vector register which memorizes a series of data, two or more preparations, It calculates to a series of data memorized by this vector register 77 with the computing element of the adder 70 of the floating point (FP:FloatingPoint), a multiplier 71, and divider 72 grade. The processing unit (the so-called supercomputer) which accelerated the operation is known by pipelining these processings.

[0007] In this processing unit, data are transmitted to the above-mentioned vector register 77 through data lines 78 and 80 with vector I/O device (vector I/O) 79 in advance of an operation from main storage 81. And the data transmitted to the vector register 77 are again written in the vector register 77 through the output bus line 74, after each above-mentioned computing elements 70-72 are supplied through the input bus line 73 and an operation is completed. Furthermore, the data written in the vector register 77 which is the result of an operation are again memorized by main storage 81 through vector I/O79.

EFFECT OF THE INVENTION

[Effect of the Invention] With the parallel operation processor applied to this invention by above-mentioned explanation so that clearly The number data of elements which a control section divides the data of the letter of a matrix into a field, and express the number of the non-zero elements which are not 0 in a field, Operation data are formed from the group of the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data. Operation data are supplied to each data-processing section, and the operation of each data-processing section is controlled. Each data-processing section The number data of elements with which each operation part was memorized by each storage section by reading location data and value data based on read-out and the number data of elements, and calculating to juxtaposition by supplying the operation data supplied from the control section to the storage section It can calculate at a high speed and the storage capacity which an operation takes can be reduced.

[0065] Moreover, in the parallel operation processor concerning this invention, by dividing for every line of the data of the letter of a matrix, forming operation data, and calculating to juxtaposition in two or more data-processing sections, a matrix can be calculated at a high speed and the storage capacity which an operation takes can be reduced.

[0066] Moreover, in the parallel operation processor concerning this invention, since the storage section was equipped with the serial access memory which performs the continuous writing and continuous read-out of data, cost performance can be raised and equipment can be miniaturized.

[0067] Moreover, in the parallel operation processor concerning this invention, a more efficient and high-speed operation can be performed by supplying operation data to each data-processing section so that the load of each data-processing section may become equal.

TECHNICAL PROBLEM

[Problem(s) to be Solved by the Invention] However, if data other than the data which read data from main storage 63 to cache memory 62 beforehand are needed, it is necessary to read the processing unit which used the above-mentioned cache memory 62 from main storage 63 (if a mistake hit is carried out). Generally, by the operation of the big matrix of size, the capacity of cache memory 62 was small, the data which have not gone into cache memory 62 needed to be read from main storage 63 each time, as for the inside of the time amount which is this read-out, CPU61 stopped and there was a problem to which the operation speed as the whole falls.

[0009] Moreover, also in the processing unit equipped with the above-mentioned vector register

77, since it was necessary to read from main storage 81 by the operation of the big matrix of size about the element (data) of the matrix which is not memorizable to the vector register 77 since the capacity of the vector register 77 is small, and it was generally necessary to calculate, operation speed was reduced.

[0010] Then, paying attention to the descriptions, such as read-out sequence of the data in the case of matrix operation, as shown in drawing 9 and drawing 10, this applicant had the serial access memory which performs the continuous writing and continuous read-out of data in each data-processing section of a parallel operation processor which calculates to juxtaposition in two or more data-processing sections, accelerated operation speed, and has proposed previously the parallel operation processor which realized low-cost-izing and a miniaturization as Japanese Patent Application No. No. 149597 [five to].

[0011] The m data-processing sections 101a and 101b which calculate as this parallel operation processor is shown in drawing 9, and 101 c*****101m, The control section 102 which outputs a control signal, operation data, etc. which control a these data-processing sections [101a-101m] operation, It has the data bus 103 which supplies the operation data from this control section 102 etc. to the above-mentioned data-processing sections 101a-101m, and the control bus 104 which supplies the control signal from the above-mentioned control section 102 to the above-mentioned data-processing sections 101a-101m.

[0012] Moreover, each above-mentioned data-processing sections 101a-101m The storage section 110 which memorizes the operation data supplied from the above-mentioned control section 102 as shown in drawing 10, The operation part 111 which calculates to the data memorized by this storage section 110, The buffer 113 which performs the communication link with the above-mentioned data bus 103 through an internal data bus 112, The buffer 115 which performs the communication link with the above-mentioned control bus 104 through the internal control bus 114, and a control signal required for a communication link are generated, and it has the control signal generating section 116 supplied to a buffer 115 and operation part 111.

[0013] Furthermore, the above-mentioned storage section 110 consists of mass continuation I/O memory (henceforth mass serial access memory (SAM)) 120 which performs the continuous writing and continuous read-out of data, and high-speed general-purpose RAM122 of capacity which performs the random writing and random read-out of the small capacity SAM 121 with a small capacity, and data and which is not so large, as shown in above-mentioned drawing 10.

[0014] In this parallel operation processor, when asking for the product of the matrix A shown, for example in the following formula 1, and Vector X, a control section 102 divides Matrix A, and assigns and supplies the data for every line to each data-processing section. For example, in the case of a formula 1, the data for two lines of 8x8 matrices are assigned to the four data-processing sections 101a-101d, and each data-processing sections 101a-101d are supplied. For example, data of the 1st line and the 5th line are supplied to data-processing section 101a, data of the 2nd line and the 6th line are supplied to data-processing section 101b, data of the 3rd line and the 7th line are supplied to data-processing section 101c, and data of the 4th line and the 8th line are supplied to 101d of data-processing sections.

[0015]

[Equation 1]

[0016] And a control section 102 is the element xi of Vector X. By supplying each data-processing sections 101a-101d, each data-processing sections 101a-101d store each line of

Matrix A in large capacity SAM 120, as shown in drawing 11, and it is the element x_i of Vector X. It stores in RAM122. And it is the element x_i of the vector X continuously corresponding to [to the time of an operation / in the each data-processing sections / 101a-101d / operation part 111] read-out and Element a_{ij} for the element a_{ij} of each line. Element y_i of the vector Y as a result of performing the operation shown in read-out and the following formula 2 from RAM122 and being based on this operation It memorizes to RAM122. And a control section 102 is the element y_i of Vector Y called for in each data-processing sections 101a-101d. Vector Y was searched for unitedly. Consequently, this parallel operation processor can perform matrix operation now at a high speed.

[0017]

[Equation 2]

[0018] On the other hand, the large-scale matrix which appears in scientific calculation was a sparse matrix most whose elements are 0 in many cases, and if all the elements of this sparse matrix were memorized at the time of an operation, there was a problem which needs huge storage capacity, so that it might be represented by a finite difference method, the finite element method, linear programming, electric circuit analysis, etc.

[0019] Then, the processing unit which shortens the operation time was considered by not performing the operation about zero element whose value of the element in this sparse matrix is 0, or not memorizing zero element.

[0020] however, in the processing unit which does not perform the operation about zero element of the above-mentioned sparse matrix the need of the positional information of the data which are not 0 in the sparse matrix used as the candidate for an operation being needed on the occasion of an operation, memorizing all elements, and judging at the time of an operation -- or It is difficult to memorize the positional information of the data which are not 0 in a sparse matrix beforehand, and to apply to the parallel operation processor which the whole operation time was increased as a result, and used the above-mentioned serial access memory by the read time of such information.

[0021] Moreover, in the processing unit which does not memorize zero element of the above-mentioned sparse matrix, there was a problem on which it is necessary to memorize the positional information of operation data in addition to operation data, the storage capacity which an operation takes increases, and the operation time as the whole increases by the read time of positional information.

[0022] This invention can be made in view of the above troubles, can calculate the matrix of big size at a high speed, can reduce the storage capacity which an operation takes, can raise cost performance, and aims at offer of the parallel operation processor which made the miniaturization possible.

MEANS

[Means for Solving the Problem] In order to solve an above-mentioned technical problem, the parallel operation processor concerning this invention The storage section which memorizes operation data, and two or more data-processing sections respectively equipped with the operation part which calculates to the operation data memorized by the storage section, The

number data of elements which divide data into a field and express the number of the non-zero elements which are not 0 in a field, Operation data are formed from the group of the location data showing the location in the field of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data. Supply operation data to each data-processing section, and it has the control section which controls the operation of each data-processing section. It is characterized by for each data-processing section reading location data and value data based on read-out and the number data of elements, calculating the number data of elements memorized in each storage section by operation part, and performing the operation to data to juxtaposition in two or more data-processing sections.

[0024] Moreover, it is characterized by for a control section dividing the data of the letter of a matrix for every line, and for the parallel operation processor concerning this invention using the number of non-zero elements in one line as the number data of elements, forming operation data per line of the data of the letter of a matrix by using the location within one line of a non-zero element as location data, and supplying them to each data-processing section.

[0025] Moreover, the parallel operation processor concerning this invention is characterized by equipping the storage section with the serial access memory which performs the continuous writing and continuous read-out of data.

[0026] Moreover, the parallel operation processor concerning this invention is characterized by a control section supplying operation data to each data-processing section so that the load of each data-processing section may become equal.

OPERATION

[Function] A control section divides the data of the letter of a matrix into a field, forms operation data from the group of the number data of elements showing the number of the non-zero elements which are not 0 in a field, the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data, supplies operation data to each data-processing section, and controls the operation of each data-processing section by the parallel-operation processor concerning this invention. Each data-processing section supplies the operation data supplied from the control section to the storage section, and operation data are memorized by the storage section. And each data-processing section reads location data and value data based on read-out and the number data of elements, and calculates to juxtaposition the number data of elements memorized by the storage section.

[0028] Moreover, with the parallel operation processor concerning this invention, it is a control section, Divide the data of the letter of a matrix for every line, and the number of non-zero elements in one line is used as the above-mentioned number data of elements. The location within the one above-mentioned line of a non-zero element is used as the above-mentioned location data, the above-mentioned operation data are formed per line of the data of the letter of a matrix from the group of the value data showing the value of the non-zero element of the location corresponding to a series of location data, operation data are supplied to the data-processing section, and the operation of each data-processing section is controlled. Each data-processing section supplies the operation data supplied from the control section to the storage section, and operation data are memorized by the storage section. And each data-processing section reads location data and value data based on read-out and the number data of elements,

and calculates to juxtaposition the number data of elements memorized in the storage section.

[0029] Moreover, a control section divides data into a field, forms operation data from the group of the number data of elements showing the number of the non-zero elements which are not 0 in a field, the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data, supplies operation data to each data-processing section, and controls the operation of each data-processing section by the parallel-operation processor concerning this invention. Each data-processing section supplies continuously the operation data supplied from the control section to serial access memory, and operation data are memorized by serial access memory. And each data-processing section reads the number data of elements memorized to serial access memory, reads location data and value data continuously based on the number data of elements, and calculates to juxtaposition.

[0030] Moreover, with the parallel operation processor concerning this invention, it is a control section, Data divide into a field, operation data form from the group of the number data of elements showing the number of the non-zero elements which are not 0 in a field, the location data showing the location of a non-zero element, and the value data showing the value of the non-zero element of the location corresponding to a series of location data, operation data supply to each data-processing section so that the load of each data-processing section may become equal, and the operation of each data-processing section controls. Each data-processing section supplies the operation data supplied from the control section to the storage section, and operation data are memorized by the storage section. And from the storage section, each data-processing section reads location data and value data based on read-out and the number data of elements, calculates by operation part, and calculates the number data of elements to juxtaposition.

EXAMPLE

[Example] Hereafter, the suitable example of the parallel operation processor concerning this invention is explained to a detail, referring to a drawing.

[0032] This parallel operation processor For example, the m data-processing sections 1a and 1b which calculate as shown in drawing 1 and 1 c*****1m, The control section 2 which outputs a control signal, operation data, etc. which control these operation control sections 1a-1m, It has the data bus 3 which supplies the operation data from this control section 2 to the above-mentioned operation control sections 1a-1m, and the control bus 4 which supplies the control signal from the above-mentioned control section 2 to the above-mentioned operation control sections 1a-1m.

[0033] Each above-mentioned operation control sections 1a-1m For example, the storage section 10 which memorizes the operation data supplied from the above-mentioned control section 2 as shown in drawing 2 , The operation part 11 which calculates to the data memorized by this storage section 10, The buffer 13 which performs the communication link with the above-mentioned data bus 3 through an internal data bus 12, the buffer 15 which performs the communication link with the above-mentioned control bus 4 through the internal control bus 14, and a control signal required for a communication link are generated, and it has the control signal generating section 16 supplied to a buffer 15 and operation part 11. Each data-processing sections 1a-1m calculate according to the control signal from a control section 2.

[0034] And the above-mentioned storage section consists of mass continuation I/O memory

(henceforth mass serial access memory (SAM)) 20 which performs the continuous writing and continuous read-out of data, and high-speed general-purpose RAM 22 of capacity which performs the random writing and random read-out of the small capacity SAM 21 with a small capacity, and data and which is not so large, as shown in above-mentioned drawing 2.

[0035] Moreover, above-mentioned size ** SAM 20 and the small capacity SAM 21 For example, the memory cell array 40 to which the writing and read-out of data are performed as shown in drawing 3, The read-out line address counter 41 which generates the read-out address of the data to the line of this memory cell array 40, The write-in line address counter 42 which generates the write-in address of the data to the line of the above-mentioned memory cell array 40, The read-out train address counter 43 which generates the read-out address of the data to the train of the above-mentioned memory cell array 40, It has the write-in train address counter 44 which generates the write-in address of the data to the train of the above-mentioned memory cell array 40, data input Rhine 45 which inputs input data Din, and data output Rhine 46 which outputs output-data Dout.

[0036] And at the time of the writing of data, the reset light signal RSTW is supplied to the above-mentioned write-in line address counter 42 and the above-mentioned write-in address train counter 44 with the write enable signal WE, a counter value is reset, the write-in clock WCK supplied henceforth is counted, and the address which writes in is generated. Furthermore, with the above-mentioned write-in clock WCK, the write-in data Din are continuously supplied from data input Rhine 45, and are memorized by the memory cell 40.

[0037] Moreover, at the time of read-out of data, the reset lead signal RSTR is supplied to the above-mentioned read-out line address counter 41 and the above-mentioned read-out address train counter 43 with the lead enable signal RE, a counter value is reset, the read-out clock RCK supplied henceforth is counted, and the address which performs read-out is generated. Furthermore, with the above-mentioned read-out clock RCK, the read-out data Din are continuously read from a memory cell 40, and it is outputted to data output Rhine 46.

[0038] Next the operation which asks this parallel operation processor actuation for the product of a matrix and a vector is made into an example, and it explains. Generally, the product Y of the matrix A as shown in the following formulas 3, and Vector X is the element yi of the product vector Y by the operation as shown in the following formula 4. It asks by asking.

[0039]

[Equation 3]

[0040]

[Equation 4]

[0041] Here, as Matrix A shows the operation of the above-mentioned formula 3 in the following formulas 5, when the element of the great portion of matrix is the so-called sparse matrix which is 0, it can ask for the product Y of Matrix A and Vector X by the operation as shown in the following formula 6.1 - a formula 6.8.

[0042]

[Equation 5]

[0043]

[Equation 6]

[0044] So, with this parallel operation equipment, a control section 2 divides Matrix A into each line, forms operation data from the group of the number data of elements showing the number of the non-zero elements in one line, the location data (row number) about all non-zero elements, and value data (value of an element), and supplies this operation data to each data-processing sections 1a-1m. In the case of the above-mentioned formula 5, since the non-example elements of the 1st line are a11 and a12, the number data of elements are set to "2", and, specifically, location data are set to "1" and "2", respectively, for example. Moreover, since the non-example elements of the 2nd line are a21, a22, and a24, the number data of elements are set to "3", location data are set to "1", "2", and "4", respectively, and value data are set to "a21", "a22", and "a24", respectively, for example.

[0045] Moreover, a control section 2 compares the number of the non-zero elements of each line, and it assigns each data-processing sections [1a-1m] operation data so that an each data-processing sections [1a-1m] load may become equal. Allocation of concrete operation data memorizes the number of the non-zero elements of each line of for example, the matrix A, and is performed by choosing the line assigned to each data-processing sections 1a-1m so that the number of these non-zero elements may become equal. For example, in the case of the above-mentioned formula 5, since the number of the non-zero elements from the 1st line of Matrix A to the 8th line is 2, 3, 1, 2, 2, 1, 3, and 2, respectively, as shown, for example in drawing 4 In the case where it has the four data-processing sections 1a, 1b, 1c, and 1d Data of the 1st line and the 4th line are supplied to data-processing section 1a, data of the 2nd line and the 3rd line are supplied to data-processing section 1b, data of the 5th line and the 8th line are supplied to data-processing section 1c, and data of the 6th line and the 7th line are supplied to 1d of data-processing sections.

[0046] Consequently, the number of the non-zero elements which the operation data for two lines are respectively supplied to each data-processing sections 1a-1d, and are supplied to each data-processing sections 1a-1d is four pieces respectively, and the each data-processing sections [1a-1d] load (operation) is equal.

[0047] And the each processing units [1a-1d] operation part 11 calculates by reading the above-mentioned operation data (the number data of elements, location data, value data) memorized to large capacity SAM 20 according to the control signal from a control section 2.

[0048] the next -- each data-processing section 1 -- the case where the product of the matrix A which shows actuation of the actual operation data in a-1d in the above-mentioned formula 3, and Vector X is calculated is explained to an example using the flow chart shown in 5 Fig. First, as operation data are supplied to each data-processing sections 1a-1d in advance of the operation, for example, it is shown in drawing 4 , the operation data for every line of Matrix A are memorized by large capacity SAM 20, and the data of Vector X are memorized by high-speed general-purpose RAM22. Specifically to the large capacity SAM 20 of data-processing section 1a The number data b1 of elements showing the number of the non-zero elements of the 1st line of Matrix A (2), The number data b2 of elements which 2 sets of location data cp (1 p= 2) and the value data dp (1 11 a 2a 12) are memorized continuously, then express the number of the non-zero elements of the 4th line of Matrix A (2), 2 sets of location data cp (2 p= 4) and the value data dp (2 42 a 4a 44) are memorized. Furthermore, the data xq (q= 1, 2...8) of Vector X

are memorized by high-speed general-purpose RAM22. Moreover, although not illustrated, the data in which the line count (the line count in their duty) and line number (i) which were assigned to each data-processing sections 1a-1d with operation data are shown are supplied to each data-processing sections 1a-1d.

[0049] And the operation data memorized by large capacity SAM 20 as mentioned above are continuously read in the sequence memorized at the time of an operation. For example, at above-mentioned data-processing section 1a, it is the number data b1 of elements of the 1st line of Matrix A first. (2) is read. Then, 2 sets of location data cp The value data dp (1 11 a 2a 12) are read continuously, and it is the number data d2 of elements of the 4th line of Matrix A further. (2) and 2 sets of location data cp The value data dp (2 42 a 4a 44) are read.

[0050] And a control section 2 supplies the control signal which starts an operation to each data-processing sections 1a-1d through a control bus 4 and the internal control bus 14, and progresses to step S1. And in step S1, operation part 11 sets to 1 the count variable k which counts a processing line count, and progresses to step S2.

[0051] In step S2, operation part 11 compares the value and the line count in its duty of the count variable k, if the count variable k is below the line count in its duty, it will progress to step S3, and if the count variable k is size from the line count in its duty, since processing for the line count in its duty is already completed, it will be ended.

[0052] In step S3, operation part 11 reads from large capacity SAM 20, the several m b, i.e., above-mentioned number data of elements, of a non-zero element, and progresses to step S4.

[0053] In step S4, the value of the count variable L is set to 1, it sets the value of Variable yi to 0, and operation part 11 progresses to step S5.

[0054] In step S5, operation part 11 measures several m of the value of the count variable L, and a non-zero element, if the value of the count variable L is several m or less of a non-zero element, it will progress to step S6, and if the value of the count variable L is size from several m of a non-zero element, since processing for one line is already completed, it will progress to step 11.

[0055] It sets to step S6 and operation part 11 is the row number cp of a non-zero element (j), i.e., above-mentioned location data, from large capacity SAM 20. It reads and progresses to step S7.

[0056] Setting to step S7, operation part 11 is, the value (aij) dp, i.e., the above-mentioned value data, of an element of the matrix from large capacity SAM 20. It reads and progresses to step S8.

[0057] It is the value xj of the element of the vector X on step S8 and corresponding to the row number (j) of high-speed general-purpose RAM22 to the above-mentioned non-zero element in operation part 11. It reads and progresses to step S9.

[0058] Setting to step S9, operation part 11 is Variable yi. The value (aij) of the element of a matrix, and value xj of the element of Vector X A product is added and it progresses to step 10.

[0059] In step S10, operation part 11 adds 1 to the value of the count variable L, and returns to step S5. That is, processing from the above-mentioned step S5 to step S10 is repeated about the data for one line, and they are the value (aij) of the element for one line, and the value xj of the element of Vector X. It is Variable yi about a product. It will add.

[0060] After processing for one line is completed, the value of the count variable L serves as size from several m of a non-zero element, and it progresses to step S11 from step S5, and sets to step S11. And operation part 11 Variable yi called for as mentioned above Value yi, i.e., the value of the element of the product vector Y, In step S12 which writes in and follows high-speed general-purpose RAM22, 1 is added to the value of the count variable k, and it returns to step S2. That is,

processing from step S2 to step S12 is repeated until processing for the line count in its duty is completed.

[0061] As mentioned above, each data-processing sections 1a-1d are the values y_i of the assigned element of the product vector Y according to the above-mentioned formula 6.1 - a formula 6.8. As it calculates, for example, is shown in drawing 6, it memorizes to high-speed general-purpose RAM22. And value y_i of the element of this product vector Y It is read by the control signal from a control section 2, and a control section 2 is the value y_i of the element of the product vector Y from each data-processing sections 1a-1d. An operation is ended unitedly.

[0062] Since this invention was applied to the parallel operation processor which performs the operation which asks for the product of a matrix and a vector, while a value cannot perform the operation to the zero element which is 0 but can accelerate an operation, in this parallel operation processor, the storage capacity of the storage section which an operation takes can be reduced, so that clearly from the above explanation. Moreover, since operation data were assigned to each data-processing section so that the load of each data-processing section might become equal as mentioned above, operation effectiveness can be raised and the whole operation time can be shortened.

[0063] In addition, the technical thought of this invention is not limited to an above-mentioned example, and can also apply the operation to process also to the operation which asks for an above-mentioned matrix, not only the product of a vector but a matrix, and the product of matrices. Moreover, the field which divides data can also perform the operation which asks for a matrix and the product of matrices at a high speed by dividing one matrix for every line, dividing the matrix of another side for every train, and forming operation data by the operation which asks not only for the line unit of an above-mentioned matrix but for an above-mentioned matrix and the above-mentioned product of matrices. Moreover, it is clear that it is applicable also to the operation to the data with which the data which serve as a candidate for an operation, for example are not restricted to an above-mentioned matrix, either, and a single string continues.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram showing the configuration of the parallel operation processor which applied this invention.

[Drawing 2] It is the block diagram showing the concrete configuration of the data-processing section which constitutes the above-mentioned parallel operation processor.

[Drawing 3] It is the block diagram showing the concrete configuration of SAM which constitutes the storage section of the above-mentioned data-processing section.

[Drawing 4] It is drawing for explaining actuation of the above-mentioned parallel operation processor.

[Drawing 5] It is a flow chart for explaining actuation of the above-mentioned train processing unit.

[Drawing 6] It is drawing for explaining actuation of the above-mentioned parallel operation processor.

[Drawing 7] It is the block diagram showing the configuration of the conventional processing unit.

[Drawing 8] It is the block diagram showing the configuration of the conventional processing

unit.

[Drawing 9] It is the block diagram showing the configuration of the conventional parallel operation processor.

[Drawing 10] It is the block diagram showing the concrete configuration of the data-processing section which constitutes the conventional parallel operation processor.

[Drawing 11] It is drawing for explaining actuation of the conventional parallel operation processor.

- 1 Data-processing section
- 2 Control section
- 3 Data bus
- 4 Control bus
- 10 Storage section
- 11 Operation part
- 12 Internal data bus
- 13 15 ... Buffer
- 14 Internal control bus
- 16 Control signal generating section
- 20 Large capacity SAM
- 21 Small capacity SAM
- 22 High-speed general-purpose RAM
- b The number data of elements
- c Location data
- d Value data
- xj Vector data
- yi Vector

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-239843

(43) 公開日 平成7年(1995)9月12日

(51) Int.Cl.⁶

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 17/16

15/16

3 9 0 Z

G 0 6 F 15/ 347

P

審査請求 未請求 請求項の数 4 O L (全 11 頁)

(21) 出願番号 特願平6-28110

(22) 出願日 平成6年(1994)2月25日

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 加藤 泰信

東京都品川区北品川6丁目7番35号 ソニ

ー株式会社内

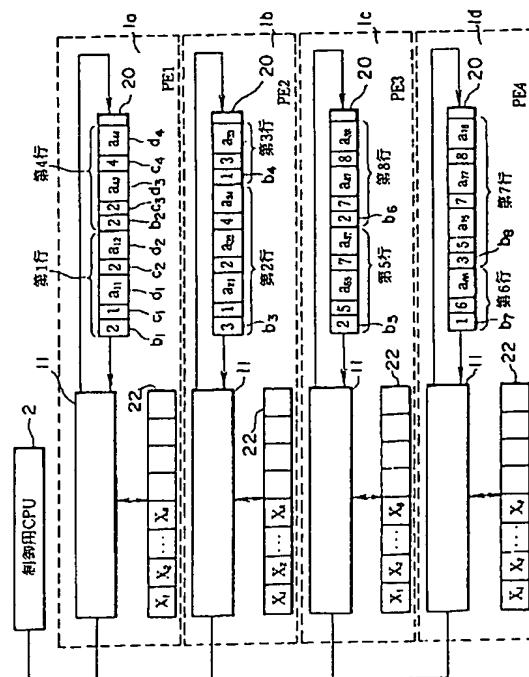
(74) 代理人 弁理士 小池 晃 (外2名)

(54) 【発明の名称】 並列演算処理装置

(57) 【要約】

【構成】制御部2は、行列Aを1行毎に分割し、1行内の非零要素の数を表わす要素数データ b_i と、すべての非零要素についての列番号 c_i と、要素の値 d_i の組から1行毎の演算データを形成し、演算データとベクトルXのデータ x_i を各演算処理部1a～1dに供給し、各演算処理部1a～1dは、演算データをSAM20に記憶し、ベクトルXのデータ x_i をRAM22に記憶する。各演算処理部1a～1dはSAM20に記憶した要素数データ b_i を読み出し、要素数データ b_i に基づいて列番号 c_i と要素の値 d_i を読み出し、列番号 c_i に対応するベクトルXのデータ x_i をRAM22から読み出して演算部11により演算を行ない、行列AとベクトルXの内積を求める演算を各演算処理部1a～1dで並列に行なう。

【効果】演算に要する記憶容量を低減でき、より効率的で高速な演算を行なうことができる。



【特許請求の範囲】

【請求項 1】 演算データを記憶する記憶部と、該記憶部に記憶された演算データに対して演算を行なう演算部を各々備えた複数の演算処理部と、

データを領域に分割し、該領域内の 0 でない非零要素の数を表わす要素数データと、上記非零要素の上記領域内の位置を表わす位置データと、一連の位置データに対応する位置の非零要素の値を表わす値データの組から演算データを形成し、該演算データを各演算処理部に供給し、各演算処理部の演算を制御する制御部とを有し、上記各演算処理部は上記各記憶部に記憶した上記要素数データを読み出し、該要素数データに基づいて上記位置データ及び値データを読み出して演算部により演算を行ない、データに対する演算を上記複数の演算処理部で並列に行なうことを特徴とする並列演算処理装置。

【請求項 2】 上記制御部は、行列状のデータを 1 行毎に分割し、該 1 行内の非零要素数を上記要素数データとし、非零要素の上記 1 行内での位置を上記位置データとして、行列状のデータの行単位で上記演算データを形成し、各演算処理部に供給することを特徴とする請求項 1 記載の並列演算処理装置。

【請求項 3】 上記記憶部は、データの連続的な書込み及び読出しを行なうシリアルアクセスメモリを備えることを特徴とする請求項 1 又は請求項 2 記載の並列演算処理装置。

【請求項 4】 上記制御部は、各演算処理部の負荷が均等となるように演算データを各演算処理部に供給することを特徴とする請求項 1 乃至請求項 3 のいずれか 1 項に記載の並列演算処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、例えば、複数の演算部によりデータの演算を並列に行なう並列演算処理装置に関し、特に、行列状のデータを演算する並列演算処理装置に関する。

【0002】

【従来の技術】 近年のコンピュータの発達と普及によって、以前は実現不可能と思われていた大規模な科学技術計算も現実的なものとなってきた。例えば電磁界解析や流体の流動解析に用いられている差分法、有限要素法、境界要素法等では、それらの解析の課程で非常に大きなサイズの連立一次方程式の求解や行列の固有値計算など、いわゆる行列演算が頻繁に現れる。

【0003】 また、解析精度を向上させるために、行列を大きくすることが考えられるが、行列が大きくなると、行列演算は、高速演算が可能なスーパーコンピュータ等をもってしても非常に時間がかかり、高速処理が望まれている。

【0004】 以上のような大きなサイズの行列演算では演算処理部の処理速度の高速化と共に、高速で記憶容量

の大きい記憶装置が求められるが、一般に記憶装置の記憶容量と動作速度は相反する関係にあるために、高速で記憶容量の大きい記憶装置は、非常に高価で、また、装置が大きくなるものとなる問題があった。

【0005】 このため、例えば図 7 に示すように、演算処理部 (CPU) 61 と大容量の主記憶装置 63 との間に高速ではあるが比較的容量の小さいキャッシュメモリ 62 を備え、予めデータを主記憶装置 63 からバスライン 65 を介してキャッシュメモリ 62 に読出しておき、CPU 61 は、バスライン 64 を介してキャッシュメモリ 62 からデータを読み出すことにより、データの高速読出しを実現し、演算を高速化した演算処理装置が使用されている。

【0006】 また、例えば図 8 に示すように、行列演算等の一連のデータに対しての同じ演算を高速に行なうために、ベクトル、すなわち一連のデータを記憶するベクトルレジスタ 77 を複数備え、このベクトルレジスタ 77 に記憶された一連のデータに対して浮動小数点 (Floating Point) の加算器 70、乗算器 71、除算器 72 等の演算器により演算を行い、これらの処理をパイプライン化することにより演算を高速化した演算処理装置 (所謂スーパーコンピュータ) が知られている。

【0007】 この演算処理装置では、演算に先だって、ベクトル入出力装置 (ベクトル I/O) 79 により、主記憶装置 81 からデータライン 78、80 を介してデータを上記ベクトルレジスタ 77 に転送する。そして、ベクトルレジスタ 77 に転送されたデータは、入力バスライン 73 を介して上記各演算器 70~72 に供給され、演算が終了した後、出力バスライン 74 を介して再度ベクトルレジスタ 77 に書き込まれる。さらに、演算結果であるベクトルレジスタ 77 に書き込まれたデータは、再度ベクトル I/O 79 を介して主記憶装置 81 に記憶される。

【0008】

【発明が解決しようとする課題】 しかしながら、上記キャッシュメモリ 62 を使用した演算処理装置は、予めデータを主記憶装置 63 からキャッシュメモリ 62 に読出したデータ以外のデータが必要になると (ミスヒットすると) 主記憶装置 63 から読み出す必要がある。一般にキャッシュメモリ 62 の容量は小さく、大きなサイズの行列の演算では、キャッシュメモリ 62 に入りきらないデータは、その都度、主記憶装置 63 から読み出す必要があり、この読出しの時間中は CPU 61 が停止し、全体としての演算速度が低下する問題があった。

【0009】 また、上記ベクトルレジスタ 77 を備えた演算処理装置においても、一般にベクトルレジスタ 77 の容量が小さいため、大きなサイズの行列の演算では、ベクトルレジスタ 77 に記憶できない行列の要素 (データ) については主記憶装置 81 から読み出して演算する

必要があるため、演算速度を低下させていた。

【0010】そこで、本件出願人は行列演算の際のデータの読出し順序等の特徴に着目し、図9及び図10に示すように、複数の演算処理部で並列に演算を行なう並列演算処理装置の各演算処理部にデータの連続的な書込み及び読出しを行なうシリアルアクセスメモリを備え、演算速度を高速化し、低コスト化及び小型化を実現した並列演算処理装置を、例えば特願平5-149597号として、先に提案している。

【0011】この並列演算処理装置は、図9に示すように、演算を行なうm個の演算処理部101a、101b、101c・・・101mと、該演算処理部101a～101mの演算を制御する制御信号及び演算データ等を出力する制御部102と、該制御部102からの演算データ等を上記演算処理部101a～101mに供給するデータバス103と、上記制御部102からの制御信号を上記演算処理部101a～101mに供給する制御バス104とを備えている。

【0012】また、上記各演算処理部101a～101mは、図10に示すように、上記制御部102から供給される演算データを記憶する記憶部110と、該記憶部110に記憶されたデータに対して演算を行なう演算部111と、内部データバス112を介して上記データバス103との通信を行なうバッファ113と、内部制御*

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} \\ a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} \quad \cdots \text{式1}$$

【0016】そして、制御部102は、ベクトルXの要素 x_i を、各演算処理部101a～101dに供給し、各演算処理部101a～101dは、例えば図11に示すように、行列Aの各行を大容量SAM120に格納し、ベクトルXの要素 x_i をRAM122に格納する。そして、演算時に各演算処理部101a～101dの演算部111は、各行の要素 a_{ij} を連続的に読出し、要素 a_{ij} に対応するベクトルXの要素 x_i をRAM122から読出し、次の式2に示す演算を行ない、この演算による結*

$$y_j = \sum_{i=1}^n a_{ij} x_i \quad \cdots \text{式2}$$

【0018】一方、差分法、有限要素法、線形計画法、回路解析等に代表されるように、科学技術計算で現れる大規模行列が要素の大部分が0である疎行列である場合が多く、演算時に、この疎行列の全ての要素を記憶するならば膨大な記憶容量を必要とする問題があった。

【0019】そこで、この疎行列内の要素の値が0である0要素についての演算を行なわない、あるいは、0要

*バス114を介して上記制御バス104との通信を行なうバッファ115と、通信に必要な制御信号を発生し、バッファ115及び演算部111に供給する制御信号発生部116とを備える。

【0013】さらに、上記記憶部110は、例えば上述の図10に示すように、データの連続的な書込み及び読出しを行なう、大容量の連続1/Oメモリ（以下、大容量シリアルアクセスメモリ（SAM）という）120と、容量の小さい小容量SAM121と、データのランダムな書込み及び読出しを行なう、容量のそれほど大きくない高速汎用RAM122とからなる。

【0014】この並列演算処理装置では、例えば次の式1に示す行列AとベクトルXの積を求めるときは、制御部102は、行列Aを分割して、各行毎のデータを各演算処理部に割り当てて供給する。例えば式1の場合では8×8行列の2行分のデータを4つの演算処理部101a～101dに割り当てて各演算処理部101a～101dに供給する。例えば演算処理部101aに第1行及び第5行のデータを供給し、演算処理部101bに第2行及び第6行のデータを供給し、演算処理部101cに第3行及び第7行のデータを供給し、演算処理部101dに第4行及び第8行のデータを供給する。

【0015】

【数1】

※果、すなわちベクトルYの要素 y_j をRAM122に記憶する。そして制御部102は、各演算処理部101a～101dで求められたベクトルYの要素 y_j を結合してベクトルYを求めている。この結果、この並列演算処理装置は、行列演算を高速に実行できるようになっていた。

【0017】

【数2】

素の記憶を行なわないことにより演算時間を短縮する演算処理装置が考えられていた。

【0020】しかしながら、上記疎行列の0要素についての演算を行なわない演算処理装置では、演算に際して、演算対象となる疎行列内の0でないデータの位置情報が必要となり、全要素を記憶しておき、演算時に判断する必要、若しくは、予め疎行列内の0でないデータの

位置情報を記憶しておく必要があり、これらの情報の読出し時間により、結果的に全体の演算時間を増大させ、上記シリアルアクセスメモリを使用した並列演算処理装置に適用することは難しい。

【0021】また、上記疎行列の0要素の記憶を行わない演算処理装置では、演算データ以外に演算データの位置情報を記憶する必要があり、演算に要する記憶容量が増大し、位置情報の読出し時間により全体としての演算時間が増大する問題があった。

【0022】本発明は、上述のような問題点を鑑みてなされたものであり、大きなサイズの行列の演算を高速に行なうことができ、演算に要する記憶容量を低減でき、コストパフォーマンスを向上させることができ、小型化を可能とした並列演算処理装置の提供を目的とする。

【0023】

【課題を解決するための手段】上述の課題を解決するために、本発明に係る並列演算処理装置は、演算データを記憶する記憶部と、記憶部に記憶された演算データに対して演算を行なう演算部を各々備えた複数の演算処理部と、データを領域に分割し、領域内の0でない非零要素の数を表わす要素数データと、非零要素の領域内の位置を表わす位置データと、一連の位置データに対応する位置の非零要素の値を表わす値データの組から演算データを形成し、演算データを各演算処理部に供給し、各演算処理部の演算を制御する制御部とを有し、各演算処理部は各記憶部に記憶した要素数データを読み出し、要素数データに基づいて位置データ及び値データを読み出して演算部により演算を行ない、データに対する演算を複数の演算処理部で並列に行なうことを特徴とする。

【0024】また、本発明に係る並列演算処理装置は、制御部が行列状のデータを1行毎に分割し、1行内の非零要素数を要素数データとし、非零要素の1行内での位置を位置データとして、行列状のデータの行単位で演算データを形成し、各演算処理部に供給することを特徴とする。

【0025】また、本発明に係る並列演算処理装置は、記憶部がデータの連続的な書込み及び読出しを行なうシリアルアクセスメモリを備えることを特徴とする。

【0026】また、本発明に係る並列演算処理装置は、制御部が各演算処理部の負荷が均等となるように演算データを各演算処理部に供給することを特徴とする。

【0027】

【作用】本発明に係る並列演算処理装置では、制御部は、行列状のデータを領域に分割し、領域内の0でない非零要素の数を表わす要素数データと、非零要素の位置を表わす位置データと、一連の位置データに対応する位置の非零要素の値を表わす値データの組から演算データを形成し、演算データを各演算処理部に供給し、各演算処理部の演算を制御する。各演算処理部は制御部から供給された演算データを記憶部に供給し、演算データが記

憶部に記憶される。そして、各演算処理部は記憶部に記憶された要素数データを読み出し、要素数データに基づいて位置データ及び値データを読み出して並列に演算を行なう。

【0028】また、本発明に係る並列演算処理装置では、制御部は、行列状のデータを1行毎に分割し、1行内の非零要素数を上記要素数データとし、非零要素の上記1行内での位置を上記位置データとし、一連の位置データに対応する位置の非零要素の値を表わす値データの組から行列状のデータの行単位で上記演算データを形成し、演算データを演算処理部に供給し、各演算処理部の演算を制御する。各演算処理部は制御部から供給された演算データを記憶部に供給し、演算データが記憶部に記憶される。そして、各演算処理部は記憶部に記憶した要素数データを読み出し、要素数データに基づいて位置データ及び値データを読み出して並列に演算を行なう。

【0029】また、本発明に係る並列演算処理装置では、制御部は、データを領域に分割し、領域内の0でない非零要素の数を表わす要素数データと、非零要素の位置を表わす位置データと、一連の位置データに対応する位置の非零要素の値を表わす値データの組から演算データを形成し、演算データを各演算処理部に供給し、各演算処理部の演算を制御する。各演算処理部は制御部から供給された演算データをシリアルアクセスメモリに連続的に供給し、演算データがシリアルアクセスメモリに記憶される。そして、各演算処理部はシリアルアクセスメモリに記憶した要素数データを読み出し、要素数データに基づいて位置データ及び値データを連続的に読み出して並列に演算を行なう。

【0030】また、本発明に係る並列演算処理装置では、制御部は、データを領域に分割し、領域内の0でない非零要素の数を表わす要素数データと、非零要素の位置を表わす位置データと、一連の位置データに対応する位置の非零要素の値を表わす値データの組から演算データを形成し、各演算処理部の負荷が均等となるように演算データを各演算処理部に供給し、各演算処理部の演算を制御する。各演算処理部は制御部から供給された演算データを記憶部に供給し、演算データが記憶部に記憶される。そして、各演算処理部は記憶部より要素数データを読み出し、要素数データに基づいて位置データ及び値データを読み出して演算部により演算を行ない並列に演算を行なう。

【0031】

【実施例】以下、本発明に係る並列演算処理装置の好適な実施例を図面を参照しながら詳細に説明する。

【0032】この並列演算処理装置は、例えば図1に示すように、演算を行なうm個の演算処理部1a、1b、1c・・・1mと、該演算制御部1a～1mを制御する制御信号及び演算データ等を出力する制御部2と、該制御部2からの演算データを上記演算制御部1a～1mに

供給するデータバス3と、上記制御部2からの制御信号を上記演算制御部1a~1mに供給する制御バス4とを備えている。

【0033】上記各演算制御部1a~1mは、例えば図2に示すように、上記制御部2から供給される演算データを記憶する記憶部10と、該記憶部10に記憶されたデータに対して演算を行なう演算部11と、内部データバス12を介して上記データバス3との通信を行なうバッファ13と、内部制御バス14を介して上記制御バス4との通信を行なうバッファ15と、通信に必要な制御信号を発生し、バッファ15及び演算部11に供給する制御信号発生部16とを備える。各演算処理部1a~1mは制御部2からの制御信号に応じて演算を行なうようになっている。

【0034】そして、上記記憶部は、例えば上述の図2に示すように、データの連続的な書込み及び読出しを行なう、大容量の連続1/Oメモリ（以下、大容量シリアルアクセスメモリ（SAM）という）20と、容量の小さい小容量SAM21と、データのランダムな書込み及び読出しを行なう、容量のそれほど大きくない高速汎用RAM22とからなる。

【0035】また、上記大容SAM20及び小容量SAM21は、例えば図3に示すように、データの書込み及び読出しが行われるメモリセルアレイ40と、該メモリセルアレイ40の行に対するデータの読出しアドレスを発生する読出し行アドレスカウンタ41と、上記メモリセルアレイ40の行に対するデータの書込みアドレスを発生する書込み行アドレスカウンタ42と、上記メモリセルアレイ40の列に対するデータの読出しアドレスを発生する読出し列アドレスカウンタ43と、上記メモリ*30

*セルアレイ40の列に対するデータの書込みアドレスを発生する書込み列アドレスカウンタ44と、入力データDinを入力するデータ入力ライン45と、出力データDoutを出力するデータ出力ライン46とを備えている。

【0036】そして、データの書込み時には、ライトイネーブル信号WEと共に、上記書込み行アドレスカウンタ42及び上記書込みアドレス列カウンタ44にリセットライト信号RSTWが供給され、カウンタ値がリセットされ、以後供給される書込みクロックWCKをカウントして、書込みを行なうアドレスを発生する。さらに、上記書込みクロックWCKと共に、書込みデータDinがデータ入力ライン45から連続的に供給され、メモリセル40に記憶される。

【0037】また、データの読み出し時には、リードイネーブル信号REと共に、上記読出し行アドレスカウンタ41及び上記読出しアドレス列カウンタ43にリセットラード信号RSTRが供給され、カウンタ値がリセットされ、以後供給される読出しクロックRCKをカウントして、読出しを行なうアドレスを発生する。さらに、上記読出しクロックRCKと共に、メモリセル40から読出しデータDinが連続的に読み出され、データ出力ライン46に出力される。

【0038】つぎに、この並列演算処理装置動作を、例えば行列とベクトルの積を求める演算を例にして説明する。一般に、以下の式3に示すような行列AとベクトルXの積Yは、次の式4に示すような演算により、積ベクトルYの要素y_iを求めることによって求められる。

【0039】

【数3】

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} \\ a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} \quad \dots \text{式 3}$$

【0040】

※ ※ 【数4】

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad \dots \text{式 4}$$

【0041】ここで、上記式3の演算において行列Aが、例えば以下の式5に示すように、行列の大部分の要素が0である所謂疎行列である場合は、次の式6、1~式6、8に示すような演算で行列AとベクトルXの積Y

を求めることができる。

【0042】

【数5】

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & a_{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{42} & 0 & a_{44} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{55} & 0 & a_{57} & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{66} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{76} & 0 & a_{77} & a_{78} \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{87} & a_{88} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} \quad \dots \text{式 5}$$

【0043】

* * 【数6】

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 & \dots \text{式 6. 1} \\ y_2 &= a_{21}x_1 + a_{22}x_2 + a_{24}x_4 & \dots \text{式 6. 2} \\ y_3 &= a_{33}x_3 & \dots \text{式 6. 3} \\ y_4 &= a_{42}x_2 + a_{44}x_4 & \dots \text{式 6. 4} \\ y_5 &= a_{55}x_5 + a_{57}x_7 & \dots \text{式 6. 5} \\ y_6 &= a_{66}x_6 & \dots \text{式 6. 6} \\ y_7 &= a_{76}x_6 + a_{77}x_7 + a_{78}x_8 & \dots \text{式 6. 7} \\ y_8 &= a_{87}x_7 + a_{88}x_8 & \dots \text{式 6. 8} \end{aligned}$$

【0044】そこで、この並列演算装置では、制御部2は、行列Aを各行に分割し、1つの行内の非零要素の数を表わす要素数データと、すべての非零要素についての位置データ（列番号）と値データ（要素の値）の組から演算データを形成し、この演算データを各演算処理部1a～1mに供給する。具体的には、上記式5の場合では、例えば第1行の非零要素は a_{11} 及び a_{12} であるから、要素数データは“2”となり、位置データはそれぞれ“1”及び“2”となる。また例えば第2行の非零要素は a_{21} 、 a_{22} 及び a_{24} であるから、要素数データは“3”となり、位置データはそれぞれ“1”、“2”及び“4”となり、値データはそれぞれ a_{21} 、“ a_{22} ”及び a_{24} となる。

【0045】また、制御部2は、各行の非零要素の数を比較して、各演算処理部1a～1mの負荷が均等になるように、各演算処理部1a～1mの演算データを割り当てる。具体的な演算データの割当は、例えば行列Aの各行の非零要素の数を記憶しておき、この非零要素の数が均等となるように各演算処理部1a～1mに割り当てる行を選択することにより行なう。例えば上述の式5の場合では、行列Aの第1行から第8行までの非零要素の数はそれぞれ2、3、1、2、2、1、3、2であるから、例えば図4に示すように、4つの演算処理部1a、1b、1c、1dを備える場合では、演算処理部1aに第1行及び第4行のデータを供給し、演算処理部1bに第2行及び第3行のデータを供給し、演算処理部1cに第5行及び第8行のデータを供給し、演算処理部1dに第6行及び第7行のデータを供給する。

【0046】この結果、各演算処理部1a～1dには各々2行分の演算データが供給され、各演算処理部1a～1dに供給される非零要素の数は、各々4個となっており、各演算処理部1a～1dの負荷（演算）が均等にな

っている。

20 【0047】そして、各演算処理装置1a～1dの演算部11は、制御部2からの制御信号に応じて、大容量SAM20に記憶した、上述の演算データ（要素数データ、位置データ、値データ）を読み出して演算を行なう。

【0048】つぎに、各演算処理部1a～1dでの実際の演算データの動作を、例えば上述の式3に示す行列AとベクトルXの積を計算する場合を例に5図に示すフローチャートを用いて説明する。まず、各演算処理部1a～1dには演算に先立って演算データが供給されており、例えば図4に示すように、大容量SAM20には、行列Aの行毎の演算データが記憶されており、高速汎用RAM22にはベクトルXのデータが記憶されている。具体的には、例えば演算処理部1aの大容量SAM20には、行列Aの1行目の非零要素の数を表わす要素数データb₁(2)と、2組の位置データc₁(p=1、2)と値データd₁(1、a₁₁、2、a₁₂)が連続的に記憶され、続いて、行列Aの4行目の非零要素の数を表わす要素数データb₄(2)と、2組の位置データc₄(p=2、4)と値データd₄(2、a₄₂、4、a₄₄)が記憶されている。さらに、高速汎用RAM22にはベクトルXのデータx₁(q=1、2、…、8)が記憶されている。また、図示しないが各演算処理部1a～1dには、演算データと共に、各演算処理部1a～1dに割り当てられた行数（担当行数）及び行番号(i)を示すデータが供給されている。

【0049】そして、上述のように大容量SAM20に記憶された演算データは、演算時に記憶された順序で連続的に読み出される。例えば上述の演算処理部1aでは、まず、行列Aの1行目の要素数データb₁(2)が読み出され、続いて、2組の位置データc₁と値データd₁(1、a₁₁、2、a₁₂)が連続的に読み出され、さ

らに、行列Aの4行目の要素数データd、(2)と、2組の位置データc。と値データd。(2、a₁₁、4、a₁₁)が読み出される。

【0050】そして、制御部2は演算を開始する制御信号を制御バス4及び内部制御バス14を介して各演算処理部1a～1dに供給し、ステップS1に進む。そして、ステップS1において、演算部11は、処理行数をカウントするカウント変数kを1にしてステップS2に進む。

【0051】ステップS2において、演算部11は、カウント変数kの値と担当行数を比較し、カウント変数kが担当行数以下であればステップS3に進み、カウント変数kが担当行数より大であれば、すでに担当行数分の処理が終了しているのであるから終了する。

【0052】ステップS3において、演算部11は、大容量SAM20から非零要素の数m、すなわち上述の要素数データbを読み出してステップS4に進む。

【0053】ステップS4において、演算部11は、カウント変数Lの値を1に、変数y_iの値を0にしてステップS5に進む。

【0054】ステップS5において、演算部11は、カウント変数Lの値と非零要素の数mを比較し、カウント変数Lの値が非零要素の数m以下であればステップS6に進み、カウント変数Lの値が非零要素の数mより大であれば、すでに1行分の処理が終了しているのであるからステップ11に進む。

【0055】ステップS6において、演算部11は、大容量SAM20から非零要素の列番号(j)、すなわち上述の位置データc。を読み出してステップS7に進む。

【0056】ステップS7において、演算部11は、大容量SAM20から行列の要素の値(a_{ij})、すなわち上述の値データd。を読み出してステップS8に進む。

【0057】ステップS8において、演算部11は、高速汎用RAM22から、上記非零要素の列番号(j)に対応するベクトルXの要素の値x_jを読み出してステップS9に進む。

【0058】ステップS9において、演算部11は、変数y_iに行列の要素の値(a_{ij})とベクトルXの要素の値x_jとの積を加算してステップ10に進む。

【0059】ステップS10において、演算部11は、カウント変数Lの値に1を加算してステップS5に戻る。すなわち、1行分のデータについて上記ステップS5からステップS10までの処理を繰り返し、1行分の要素の値(a_{ij})とベクトルXの要素の値x_jとの積を変数y_iに加算することになる。

【0060】そして、1行分の処理が終了すると、カウント変数Lの値が非零要素の数mより大となり、ステップS5からステップS11に進み、ステップS11において、演算部11は、上述のように求められた変数y_i、

の値、すなわち積ベクトルYの要素の値y_iを高速汎用RAM22に書き込み、続くステップS12においてカウント変数kの値に1を加算してステップS2に戻る。すなわち、担当行数分の処理が終了するまで、ステップS2からステップS12までの処理を繰り返す。

【0061】上述のように、各演算処理部1a～1dは、例えば上記式6、1～式6、8に従って、割り当てられた積ベクトルYの要素の値y_iを計算し、例えば図6に示すように、高速汎用RAM22に記憶する。そして、この積ベクトルYの要素の値y_iは制御部2からの制御信号により読み出され、制御部2は、各演算処理部1a～1dからの積ベクトルYの要素の値y_iを結合して演算を終了する。

【0062】以上の説明から明らかなように、この並列演算処理装置では、行列とベクトルの積を求める演算を行なう並列演算処理装置に本発明を適用したから、値が0である零要素に対する演算を行なわず、演算を高速化することができると共に、演算に要する記憶部の記憶容量を低減させることができる。また、上述のように各演算処理部の負荷が均等になるように、演算データを各演算処理部に割り当てたから、演算効率を向上させることができ、また、全体の演算時間を短縮することができる。

【0063】なお、本発明の技術的思想は上述の実施例に限定されるものではなく、処理する演算は上述の行列とベクトルの積に限らず、例えば行列と行列の積を求める演算にも適用することもできる。また、データを分割する領域も上述の行列の行単位だけではなく、例えば上記の行列と行列の積を求める演算では、一方の行列を1行毎に分割し、他方の行列を1列毎に分割して演算データを形成することにより、行列と行列の積を求める演算を高速に行なうことができる。また、例えば演算対象となるデータも上述の行列に限るものではなく、一連の連続するデータに対する演算にも適用できることは明らかである。

【0064】

【発明の効果】上述の説明で明らかなように、本発明に係る並列演算処理装置では、制御部は、行列状のデータを領域に分割し、領域内の0でない非零要素の数を表わす要素数データと、非零要素の位置を表わす位置データと、一連の位置データに対応する位置の非零要素の値を表わす値データの組から演算データを形成し、演算データを各演算処理部に供給し、各演算処理部の演算を制御し、各演算処理部は、制御部から供給された演算データを記憶部に供給し、各演算部は各記憶部に記憶された要素数データを読み出し、要素数データに基づいて位置データ及び値データを読み出して並列に演算を行なうことにより、演算を高速に行なうことができ、演算に要する記憶容量を低減できる。

【0065】また、本発明に係る並列演算処理装置で

は、行列状のデータの1行毎に分割し、演算データを形成し、複数の演算処理部で並列に演算を行なうことにより、行列の演算を高速に行なうことができ、演算に要する記憶容量を低減できる。

【0066】また、本発明に係る並列演算処理装置では、記憶部がデータの連続的な書き込み及び読出しを行なうシリアルアクセスメモリを備えたため、コストパフォーマンスを向上させることができ、装置を小型化することができる。

【0067】また、本発明に係る並列演算処理装置では、各演算処理部の負荷が均等となるように演算データを各演算処理部に供給することにより、より効率的で高速な演算を行なうことができる。

【図面の簡単な説明】

【図1】本発明を適用した並列演算処理装置の構成を示すブロック図である。

【図2】上記並列演算処理装置を構成する演算処理部の具体的な構成を示すブロック図である。

【図3】上記演算処理部の記憶部を構成するSAMの具体的な構成を示すブロック図である。

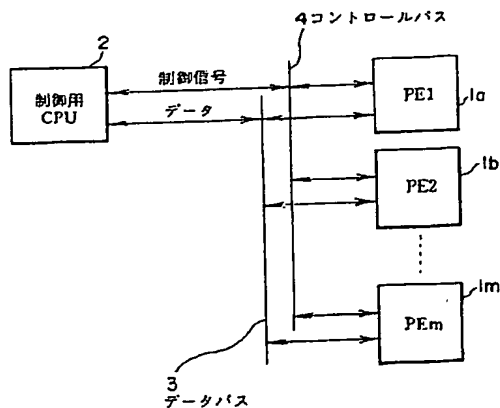
【図4】上記並列演算処理装置の動作を説明するための図である。

【図5】上記列演算処理装置の動作を説明するためのフローチャートである。

【図6】上記並列演算処理装置の動作を説明するための図である。

【図7】従来の演算処理装置の構成を示すブロック図で*

【図1】



*ある。

【図8】従来の演算処理装置の構成を示すブロック図である。

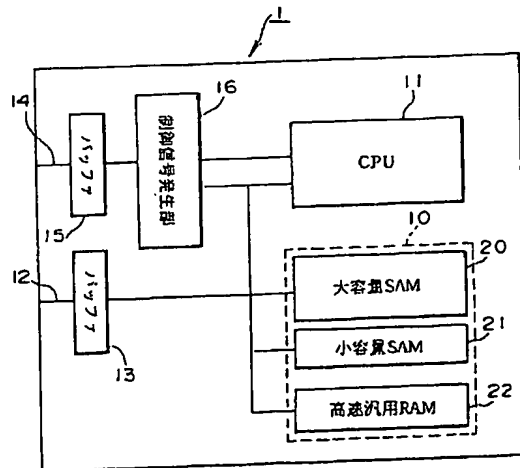
【図9】従来の並列演算処理装置の構成を示すブロック図である。

【図10】従来の並列演算処理装置を構成する演算処理部の具体的な構成を示すブロック図である。

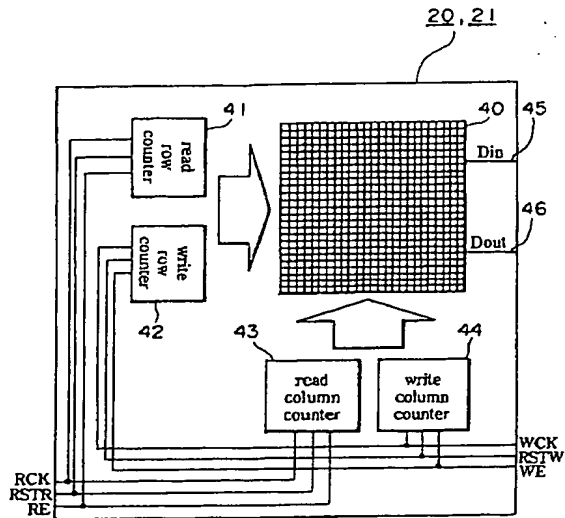
【図11】従来の並列演算処理装置の動作を説明するための図である。

- 10 演算処理部
- 2 制御部
- 3 データバス
- 4 制御バス
- 10 記憶部
- 11 演算部
- 12 内部データバス
- 13, 15 バッファ
- 14 内部制御バス
- 16 制御信号発生部
- 20 大容量SAM
- 21 小容量SAM
- 22 高速汎用RAM
- b 要素数データ
- c 位置データ
- d 値データ
- x₁ ベクトルデータ
- y₁ ベクトル

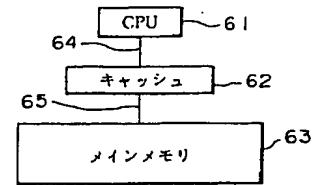
【図2】



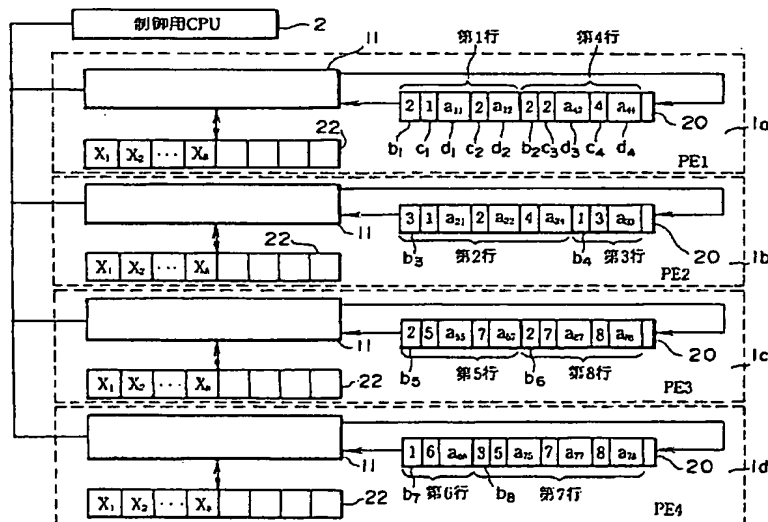
【図3】



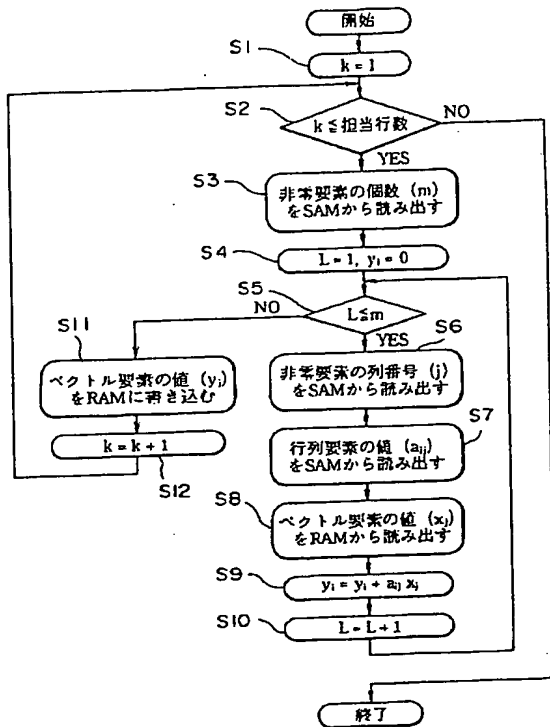
【図7】



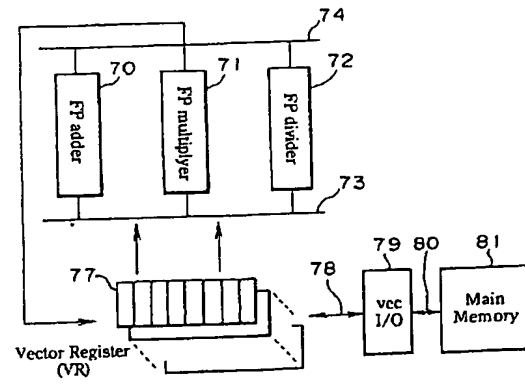
【図4】



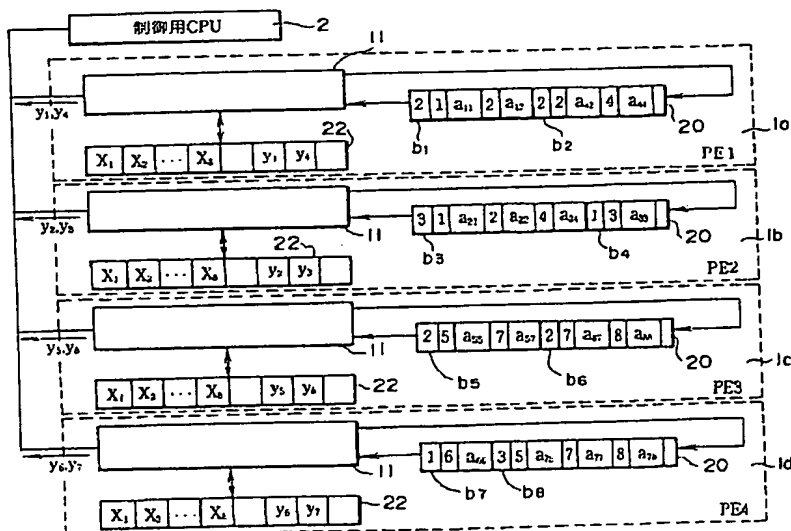
【図5】



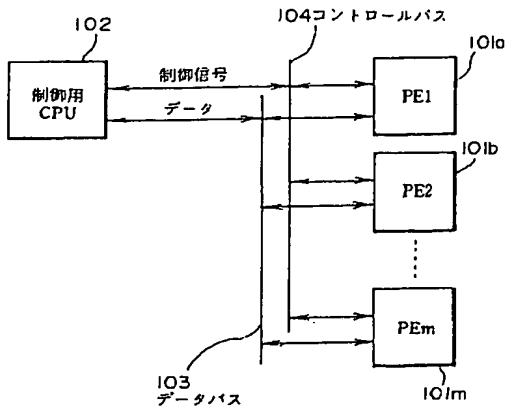
【図8】



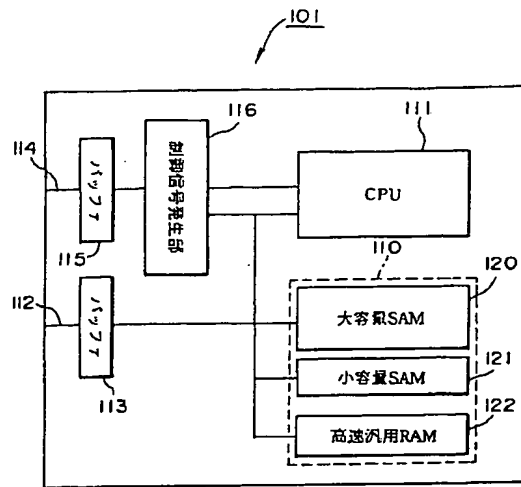
【図6】



【図9】



【図10】



【図11】

